CHAPTER 10

Proper Assignment of Accommodations to Individual Students

JENNIFER KORAN AND REBECCA KOPRIVA

Even as tests may be developed and accommodated to specifically address the needs of English language learners (ELLs), if there is no technically rigorous mechanism in place to get the specific methods to the specific students who need them, it is argued that these efforts have little effect. Several researchers who investigate accommodation effectiveness for ELLs and students with disabilities point out that consistent and appropriate accommodations decision-making is critical to the validity of standardized academic testing programs and the ability to properly use scores to compare student performance across states and districts (e.g., Kopriva *et al*., 2006a; Fuchs *et al*., 2000a; Kopriva, 2000; Hollenbeck *et al*., 1998). At the individual level when accommodations decisions are not appropriate to meet the needs of individual students, test results misrepresent their knowledge and skills (Hipolito-Delgado and Kopriva, 2006). At the aggregate level, when accommodations decisions are inconsistent from classroom to classroom or district to district, comparisons across classrooms and districts may be unfair and meaningless (Abedi, 2006; Solomon *et al*., 2001; Fuchs *et al*., 2000).

Test accommodations for ELLs are meant to remove systematic measurement error that exists while the students gain proficiency in the English language and knowledge of common U.S. cultural and educational contexts. Elliott and others (see below) have identified sixteen critical access skills that may present problems for students with disabilities when they take tests under typical standardized testing conditions. Several of these

are salient for ELLs and provide examples of the type of systematic measurement error that accommodations are designed to remediate. When choosing the accommodations a student should receive—including consideration of an option in which the student receives no test accommodations—the goal is to choose the accommodation(s) that reduce the overall amount of error due to factors irrelevant to the constructs being measured on a specific test for the student. While it is probable that other sources of error could be added whenever conditions are varied, for some students it is clear that taking the assessment under non-accommodated conditions will result in scores that substantially misrepresent the knowledge and skills of those students. The Office for Civil Rights (2000) at the U.S. Department of Education specified that the important guideline is which alternative provides better validity of scores, taking into consideration a practicable limit to cost and resource allocations that would be needed to address these issues properly.

Ideally, matching students with appropriate accommodations can be viewed as a process of selecting the most optimal package of accommodations for that student. The selection of a package of accommodations is focused on effectively reducing measurement error for a specific ELL. We will use "matching" and "accommodations decision making" interchangeably throughout this chapter.

In the large-scale assessment literature, research on the most effective accommodations for both English learners and students with disabilities is ongoing. Concurrently, movement is also being made to appropriately integrate these students in standardized testing systems designed for the broad range of students being educated in U.S. schools (No Child Left Behind act of 2001, 2002). It is generally believed that accommodations should be selected to meet the needs of the individual student. However, research confirms that one cannot validly assign accommodations to groups of students based on some broad classification or status (Sireci *et al*., 2003). Emerging work suggests that systematic methods of assignment may work better than relying on current policy approaches. Further, it presents evidence that using systematic methods to match the particular needs and strengths of individual students to specific accommodations which address these needs and strengths may increase validity and be superior to using educator directed decision-making alone (Koran *et al*., 2006; Kopriva *et al*., in press; Tindal, 2006).

Key Assignment Considerations

This section will summarize three key factors which impact the assignment process for English language learners. The first focuses on student factors which appear to be the most salient for making differential decisions about

testing options. Next is a brief summary of the primary accommodations or options which appear to relieve access barriers caused by traditional testing methods, and, finally, issues of quality, and capacity and opportunity are outlined. The latter considerations may influence the decisions, or they may influence the inferences of the scores.

Student Factors

Butler and Stevens (1997) focused on three critical background factors that they believe impact accommodation selection: English language proficiency, prior formal schooling, and length of time in the U.S. Abedi (2006) and Rivera and Collum (2006) largely focus on levels of English language proficiency and the language of instruction. In an extensive literature review, cognitive labs, and interviews and focus groups, Kopriva and others isolated relevant student factors that appear to be salient for matching ELLs to proper accommodations (e.g., Kopriva *et al*., 2007a, Kopriva *et al*., 2005a; Winter *et al*., in press). As illustrated in Figures 10.1–10.4, the variables are levels of English and home language proficiency (in reading, writing, speaking, and listening), and an aggregate variable named cultural proximity. The authors also identified key information related to teacher judgments of student needs and compensatory strengths, and teacher data related to ongoing classroom experiences with accommodations and other testing forms and formats. Cultural proximity, as Kopriva and others have defined it, is an aggregate variable which refers to the similarity between the student's native country schooling (if applicable) and selected home school-like experiences, relative to U.S. schooling and testing opportunities. Aspects of this variable are highlighted by several researchers, including Saville-Troike (1991) and Butler and Stevens (1997). Information obtained focuses on native country schooling experiences, including resources, time and pedagogical approaches, and purposes of testing. The time and consistency of the students' experience in U.S. schools is also obtained under this category as ELL experts know this is a salient and ongoing issue for this population. Selected literature associated with these aspects will be briefly explained below.

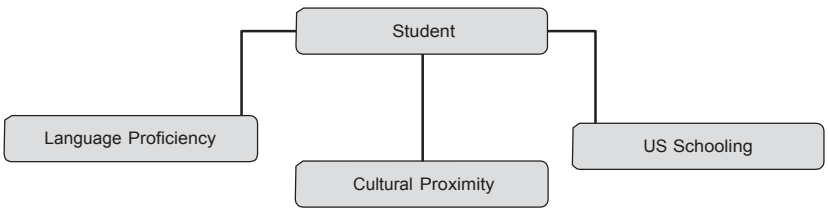

Figure 10.1 Overview of relevant student factors

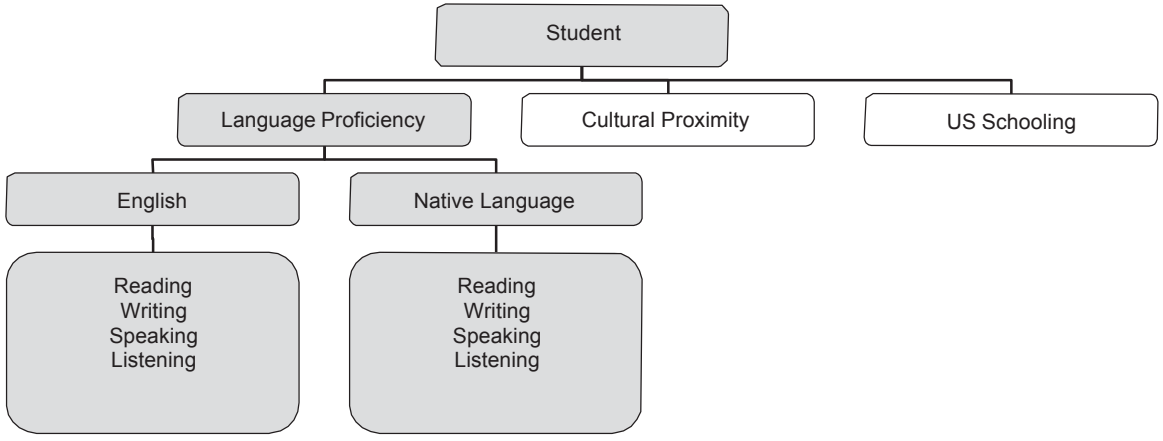

Figure 10.2 Relevant language proficiency student factors

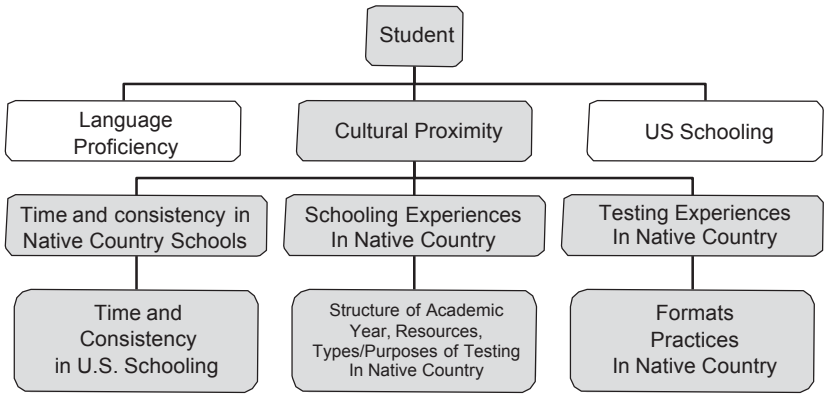

Figure 10.3 Relevant cultural proximity student factors

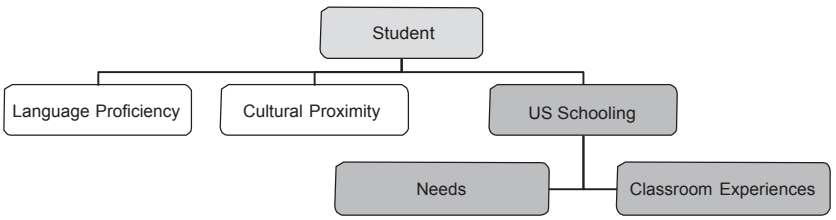

Figure 10.4 Relevant U.S. schooling student factors

The Influence of Learning English

In most large-scale testing situations, students are assumed to have reached a level of English language familiarity and reading ability necessary to understand and successfully respond to test items presented to them in a standard format. Writing proficiency in English is also sometimes key. English language learners, however, are typically not at this level in reading, writing or oral communication although they may have reached varying degrees of language proficiency in each of the domains (LaCelle-Peterson and Rivera, 1994; Durán, 1997). Clearly, ignoring the varying levels of language proficiency reached by ELLs impacts the validity of test performance for those students (Kopriva, 2000). Similarly, lumping ELLs into dichotomous "Proficient/Non-proficient" or ELL/not ELL groups does not take into account the important and salient differences in the level of proficiency each individual student may possess. In addition to limiting the usefulness of the test scores for agency use, this has been a large problem

in research associated with large scale testing accommodations (for instance see Duncan *et al*., 2005).

Considering students' proficiency in reading as opposed to listening, or writing as opposed to speaking, may have an impact on general understanding of information, learning, and test performance as well. Solano-Flores (2002) and Solano-Flores and Trumbull (2003) found that sometimes different linguistic demands are met by the students' English capabilities in the different domains. LaCelle-Peterson and Rivera (1994) emphasize that different populations of English learners differ in the process of learning English in terms of whether spoken or written expression develops more quickly. Further, as explained in Chapters 4 and 5, aspects of language knowledge appear to be somewhat differentially problematic for ELL students and native English speakers who have problems in reading or one of the other domains. ELL students present compensatory strengths as well that may or may not be the same as those of native speakers with language difficulties. Several (including Butler, Stevens, and Castellon-Wellington, 2000) emphasize that students need to learn academic English as well as conversational English and this has recently prompted a revision in English proficiency testing in order to properly encompass measurement of these crucial skills. Chapin *et al*., (2003) have identified procedures for how academic English is learned within the instruction of content instruction in English. This has direct implications for the efficacy of measurement of content knowledge, particularly the more complex academic concepts. Finally, fatigue and frustration based on language load and unfamiliar testing exposure often adds necessary testing time, and will also likely determine that students may need more frequent breaks (Durán, 1997). Because of the psychosocial factors, accommodations such as these may need to be added as well to the overall testing package for identified individual students when they are tested in English.

The Influence of the Native Language

While students may be "limited" in English proficiency, if some English proficiency can be assumed (usually reading), bilingual children often have greater overall linguistic expertise and advantage compared with their monolingual peers who are having trouble reading tests (Kester and Peña, 2002). Unequal proficiency in native language abilities may impact testing and necessitate accommodations to compensate for this strain. Research in this area suggests that using native language proficiency represents a collaborative "meaning-making" process with which to facilitate the measurement of academic content (Knapp, 1995; Ruddell, 1995). Oral L1 proficiency is particularly important if they are being instructed in L1 or if they are not yet literate enough in their native language. Additionally, oral

L1 aid for non-glossed words may be useful for students with no to little written reading proficiency in their native language and if they have been learning content in English for a short period of time (Mann *et al*., 2006).

If and as they have some degree of written native language proficiency, written aid in L1 can supplement tests in English for those in English-speaking content classrooms. This because they can get clues about English item meaning from L1 cognates and other linguistic components (August *et al*., 2004). Several studies (for example, Collier and Thomas, 2002) have shown that as students are taught literacy skills in their home language this seems to have an effect on educational outcomes such as test scores, academic performance, etc. Even when ELL students are rather literate in English, Solano-Flores, Trumbull, and Kwan (2003, April) suggest that native language aids can continue to be helpful. Finally, several authors (including Rivera and Stansfield, 2001; Duncan *et al*., 2005) emphasize that native language and dual language forms, as well as use of language aids (particularly if they are written), takes more time than if students are proficient in English and using English forms. For instance, forms in Spanish are approximately one-third longer than their counterparts in English. Knowledge about these components in ELLs' home language reveals details about their level of proficiency which offers relevant information about these students' specific test-taking needs (LaCelle-Peterson and Rivera, 1994; Durán, 1997; Butler *et al*., 2000).

The Influence of Culture

It has been argued that to better understand children's learning, one must first appreciate the cultural environment in which such learning occurs (Boykin, 1995; Greeno, 1989; Rogoff, 1990). Research indicates that students whose home culture is similar to their "school culture" (typically based on mainstream, middle-class U.S. culture, according to Farr and Turnbull (1997)) tend to fare better educationally than those whose home and school cultures differ substantially (Boykin and Bailey, 2000; Grenfell *et al*., 1998). Clearly, language minority students come from many different cultural backgrounds and life experiences that affect their educational needs (e.g., Hakuta and Beatty, 2000; LaCelle-Peterson and Rivera, 1994). While many ELLs share the common goal of improving their English language proficiency, they differ in cultural background, home language, family and educational history and so on. ELLs are from various distinct communities, and research in this area asserts that each of these communities may interact with schooling very differently (Au, 1980; Au and Jordan, 1981; Butler and Stevens, 1997; Jacob and Jordan, 1987; LaCelle and Rivera, 1994; Ogbu, 1987; Phillips, 1983; Trueba, 1989). Therefore, several of these researchers suggest it is important that information be obtained about some of the

cultural experiences of the student, especially as they relate to their schooling experiences that are key to understanding classroom and large scale testing environments.

Outcomes from this body of research also find that careful consideration should be paid to particular components of the experiences of the students' parents and the expectations of their home communities to the extent that these are different than the mainstream U.S. testing perceptions. Further, knowledge of student strengths and challenges given their home communities and cultural backgrounds seems to be salient. Malagon-Sprehn and Kopriva (2004) reported that that, because a large percentage of ELLs were native born or in home communities somewhat isolated from mainstream U.S. culture, the culture of their home communities seems to have an impact in some of the same way that experiences in home countries might and they suggest that this as well as home country experiences should be considered. Butler and Stevens (1997) report that length of time in the U.S. as well as amount and type of prior formal schooling is a necessary but not sufficient "proxy" for the degree of cultural connections and disconnections a student has with the U.S. schooling and testing culture. Additionally, Lara and August (1996) maintain that consistency in the U.S. schooling is closely related and should be considered. The time and school consistency in U.S. are often used to suggest the degree of English language proficiency as well.

The Influence of Current Schooling Experiences

Much of the literature associated with the schooling of English learners discusses that teachers assess the content and ancillary needs of their students on an ongoing basis, and individualize and adjust instruction to address them (e.g., Farr and Trumbull, 1997; O'Connor and Chapin, 2006). Because language plays such a central role in the traditional delivery and assessment of content in the classroom, these teachers have had to learn how to separate out these limits while they attempt to teach or assess other targeted knowledge. As part of this endeavor, particularly for those students with noticeably limited language skills, the teachers are trained to be attentive to the ancillary strengths of the students and utilize these characteristics on an ongoing basis. Kopriva (2005a) and Koran, Kopriva, Emick, Monroe and Garavaglia (April, 2006) have noted that teachers of ELLs appear to be able to differentially report some language needs of students, psychosocial concerns, and some compensatory strengths that students have exhibited in the classroom, when they are directly asked about these characteristics. Further, there is some evidence to support that, when teacher reports of these factors are paired with accommodations which should ameliorate the needs and support the strengths, there seems to be some boost in test performance and possibly validity of the test scores for

the students (Hipolito-Delgado and Kopriva, 2006; Kopriva *et al*., 2007). For all these reasons, it appears that this compensatory experience in classrooms of ELL students should be important to consider when accommodations are being selected, and that teachers seem to be in a position to provide information about student strengths and challenges.

While teachers of ELL students appear to be skilled at identifying student needs and strengths, and can apparently accommodate them in the classroom, there is not much evidence that they can translate their classroom rationale into reliably recommending appropriate large scale accommodations for their students (e.g., Douglas, 2004). This will be explained in detail later in this chapter, but it becomes one of primary reasons why some type of additional guidance about matching needs to be considered.

As the alignment between large scale accommodation constraints and teacher skills are being sorted out, the issue of adequate practice needs to also be considered. Most authors (e.g., Durán, 1997; LaCelle-Peterson and Rivera, 1994; Hakuta and Beatty, 2000) agree that the best methods for testing ELLs in standardized testing environments are to mirror, as much as possible, instructional and evaluation techniques which occur naturally in the classroom. Only a subset of these is reasonably feasible, however, and decisions about which large scale accommodations will be used for students will be, to some extent, externally imposed. Therefore, it remains important that every effort should be made to make sure there has been adequate use of whatever techniques are available prior to testing. The methods include, but are not limited to, compensatory techniques which might be practicable in large scale settings, but which may or may not have been used by individual teachers. They also include the practice of relevant item types or other relevant test mechanics. As noted later in this chapter, the importance of practice provides a tension somewhat between when the matching recommendations should be collected and occasions for adequate trials. Because many of the salient testing capabilities of ELLs grow over time, often rather rapidly, recommendations made later in the year will ensure that more up-to-date needs are being reflected. However, this may not give teachers the time that is needed for students to learn about and/or practice less familiar options. Because of the heavier burden on teachers of ELL students and on the students themselves to teach and learn the language plus the academic content, time for sufficient practice is in short supply. Adequate practice presents another challenge but one that is an important aspect to the appropriate assignment of accommodations.

Accommodations

Based on research findings and practitioner experiences, Rivera and Collum (2006) have recently presented the subset of accommodations that

seem to be the most salient for English learners. Pennock-Roman and Rivera (2006) have also reported effect size benefits in their meta-analyses of relevant research to date. Promising accommodations are outlined in several chapters in this book, as well, that are consistent with what these authors are reporting. As noted throughout, the research findings are at various levels of consensus on the most beneficial accommodations for this population. However, this is largely because the field has continued to learn about how to evaluate effectiveness in general (Sirici *et al*., 2003), and particularly how to determine effectiveness for students with specific needs.

Kopriva (2005a) reported that a task analysis of the active characteristics was completed for fifteen promising accommodations. Active characteristics are those aspects of the accommodations which define which skills are needed to access them. For instance, oral administration in English requires that a student have sufficient auditory capability. Further, they should be able to understand the language of English at the level of language complexity in the written text which is being recited. The fifteen accommodations were selected based on literature reviews and pilot findings in this project and another related study (Kopriva and Mislevy, 2005). For the most part, they were consistent with what was cited by Rivera and Collum, and Abedi. Once the analyses of the accommodations were completed, the options were preloaded into the assignment system which was being built. The list of the accommodations in Table 10.1 has been adapted from this work. The system will be discussed later in this chapter, but the logic essentially begins with a screen which evaluates the student's capability of taking the assessment under standard testing conditions. These conditions include a standard form as well as the testing procedures typically used in conducting large-scale testing. Therefore, accommodations are activated in the system when minimum thresholds for taking the assessment under standard conditions are not meant. Because the accommodations are discussed at length in other chapters, a discussion of the options, their research base, and their effectiveness will not be repeated here.

Kopriva (2005a) also completed a task analysis of active characteristics for two types of pretest supports, a Family Assessment Night and individually tailored classroom support. Subsequently, they were also preloaded into the assignment system pretest support is briefly discussed in Chapter 9 but essentially attends to student needs in the following ways. Family Assessment Night works with the family to understand, explain and clarify previous and current expectations and conventions of testing. Classroom support directs the teacher to a range of support which a particular student appears to need. Many of these considerations native language students take for granted, such as practice with types of items that ask questions and items

TABLE 10.1 Promising accommodations for English language learners

- Forms
  - Access-based form in English
  - Native language or dual language forms as available
- Tools
  - Bilingual word list, general or test specific
  - Picture-word dictionary
  - Problem solving tools
- Administration
  - Oral English
  - Oral home language
  - Small group
  - Language liaison
  - Extra time
  - More frequent breaks
- Response
  - Written in native language or code switching
  - Oral English
  - Oral in native language or code switching
  - Demonstrated or modeled response

that ask for student's original thinking. In all, pretest support was included because these avenues may be useful in helping to ameliorate certain cultural discontinuities of newer ELL students which, at the present time, are not adequately met even when options such as those identified in Table 10.1 are employed.

Capacity

Solano-Flores and Trumbull (Chapter 8 in this volume) argue that decisions related to the language of testing should attend to the institutional quality of L1 and English proficiency information, the institutional capacity to implement the test in L1 (text and/or oral), and their capacity to provide other accommodations. They contend that these variables should factor into the assignment decisions as well. They suggest it is not enough to produce a matching system without addressing the rigor of the information which goes into the assignment. Nor is it satisfactory to develop an assignment protocol without considering issues of agency capacity and district or state level oversight and auditing procedures which should be part of any decision system.

One important variable that has been identified but is not currently part of any matching system to date is opportunity to learn (e.g., LaCelle-Peterson and Rivera, 1994; Durán, 1997; Butler and Stevens, 1997; Hakuta

and Beatty, 2000). Other than opportunity to practice accommodations and the types of questions being used on U.S. mainstream assessments, understanding the students' learning opportunities is clearly key to properly interpreting testing inferences. While this does not impact the procedures associated with assignment of accommodations *per se*, it is a crucial aspect of accurately interpreting results based on appropriate matching decisions. It is hoped that future work will determine how to integrate this information into an effective system for assessing English language learners.

Defining the Task

Studying test accommodations and understanding how they work can be quite complicated. Tindal and Ketterlin-Geller (2004) note that in accommodations research, "few accommodations can be viewed as single, isolated variables," and that ". . . accommodations are best thought of as a package in which no one accommodation is ever studied very well in isolation" (p. 7). On one hand, test changes that we think of as a single accommodation, such as reading a test aloud to a student, may have multiple facets that remove or contribute construct irrelevant variance for a particular student. For example, if a test is read aloud to a group of students, it is possible that students may benefit from the effect of having the test administrator pace them through the test, which is different from the benefit derived by having the reading demand of the test eliminated or reduced (Hollenbeck *et al*., 2000; Weston, 2003).

Furthermore, groups of such accommodations may also be thought of as packages (Hollenbeck *et al*., 2000), as the multiple facets of various accommodations interact to produce a net increase or decrease in construct irrelevant variance. Linguistic accommodations in particular (such as different forms of the test that incorporate linguistically-oriented changes in the presentation of items or the availability of language aids, for instance a bilingual word list or decreased language complexity) function most optimally as packages. For example, side-by-side (dual language) administrations require extra time because working back and forth between the two language versions of each item takes more time than completing a comparable test in one language (Choi and McCall, 2002). As Chapter 1 discusses, this seems to be particularly true for ELL students with little English proficiency who may be especially affected by their previous schooling environments, past experiences, and current home environments, which may be very different than those associated with mainstream U.S. students.

One may think of multiple possible combinations of accommodations that function together to meet the student's need. One package of

accommodations may be considered the preferred package for a particular student although other packages compensate for the student's need nearly as well. For example, an ELL may have a level of English proficiency such that a bilingual word list may not be needed if the student is to receive a form of the test where items are presented in plain or simplified English or in their native language. However, if that student were to receive a standard English version of the test form, he or she would need the bilingual word list to help him or her access the items. In selecting accommodations for ELLs most accommodation researchers are now recognizing that the package must be tailored to the needs of the individual student. Further, over-accommodating can be just as problematic for the student as not providing the needed accommodations at all. Administering a test with all possible accommodations may be overwhelming and possibly counter-productive (Kopriva, 2005a) Administering a test with improper "bells and whistles" (particularly unnecessary additions to the test items or forms) that are not needed can be distracting (for instance, see Sharrocks-Taylor and Hargreaves, 1999).

The need to make individual accommodations decisions is common to both ELLs and students with disabilities. However, less research has been published regarding the processes used to make accommodations decisions for ELLs. In the next sections we discuss policy-based methods and research approaches. Policy methods generally reflect current practice in selecting appropriate test accommodations for individual students in both groups, although the mechanisms associated with these assignments are somewhat different. Research-based methods are emerging and approach the task from the perspective of how the literature might structure the decision process and narrow down choices for students with particular needs. Where research is lacking with respect to ELLs, we have supplemented the discussion with relevant research concerning similar issues with students with disabilities.

Policy-Based Approaches

Current guidelines for selecting accommodations for students with disabilities primarily stems from authorizations of federal legislation (Individuals with Disabilities Education Act, IDEA). Regulations or instructions for assigning accommodations to individual ELLs, on the other hand, are generally policy-based, most often at the state level. The practice for assigning large-scale accommodations for students with disabilities typically focuses on the role of the Individualized Education Plan (IEP). In addition to developing and evaluating each student's learning goals and instructional plans, the IEP addresses the proper test accommodations appropriate for each student at both the classroom and standardized testing levels. Current

practices typically used to assign large scale test accommodations to individual ELLs reflect that decisions are generally made by a single person (commonly the student's teacher or the school ELL specialist). Reviews of ELL test accommodations policies and practices both across the US and within certain states and districts confirmed that ELL teachers still are the primary decision makers when it comes to deciding ELL accommodations (Rivera and Stansfield, 2000; Kopriva *et al*., 2007a). Some educational agencies are beginning to use teams of people which are similar to IEP teams for students with disabilities (for instance, involving parents, teachers, specialists, and administrators) to make the decisions (Kopriva *et al*., 2006a). However these teams differ in major ways from IEP teams, particularly in the scope of their charge. For English learners the scope is much more constrained and standardized test focused than that defined under federal statute for students with disabilities.

Some disability researchers emphasize the importance of the IEP as a decision-making process as well as a document (Shriner and DeStefano, 2003). However, the protocol for IEP as process is often not well defined or straightforward for a team of people to implement, and similar problems may be found in the guidelines for teams making accommodations decisions for ELLs. In both situations guidelines tend to offer broad parameters rather than specific guidance for those who must make accommodations decisions. Both individual teachers and teams making accommodations decisions attempt to work within the policies given to them by the federal, state or local educational agency, but these policies generally do not contain specific recommendations for how to address the needs of specific students. Thurlow and others (e.g., 2002; 2004) routinely collect information documenting a great deal of variation and inconsistency in allowed accommodations across states for students with disabilities and ELLs. A more in-depth look at selected agencies found substantial variation in the explicitness and formality of policies for both populations among partner states and districts (Kopriva and Hedgspeth, 2005).

It appears that, presently, both ELL individuals or teams and IEP teams tend to rely heavily on the implicit and ill-defined sense of knowledge about the student obtained from the child's primary teacher(s) and/or well-meaning adults. Research on accommodations decisions for students with disabilities shows that, with little additional guidance, they tend to be inconsistent decision makers when it comes to assigning appropriate accommodations for individual students. Ysseldyke *et al*., (1994) cite vague accommodations guidelines and altruistic motivations (such as assigning an accommodation to lessen emotional distress to the student rather than because the student needs the accommodation to receive a valid score) on the part of the local decision making team as contributing to the

inconsistencies found in assigning accommodations for students with disabilities. Fuchs *et al*., (2000a) found that teacher judgments about whether to assign accommodations for students with disabilities were associated with some demographic and performance variables where there were no such associations with the accommodations decisions made by an objective accommodations decision making system.

Douglas (2004) found that ELL teachers had difficulty articulating the specific decision making process that they use in assigning accommodations. Koran and Kopriva (2006) found that teacher recommendations, unfortunately, were not statistically different from random assignment of accommodations to ELL students. This was true when teachers were asked to provide guidance without additional training from that provided by their states or districts. However, it was true even when the teachers took the leadership role themselves in a standard data collection process designed especially for this purpose. Similarly, Fuchs *et al*., (2000a) reported that teacher judgments did not correspond very well to learning disabled students' demonstrated differential boost using each of three accommodations on different alternated forms of a reading assessment. In another study, teachers predicted with no more than a chance level which special education students benefited from read aloud accommodation on a mathematics test (Helwig and Tindal, 2003). Further, Weston (2003) found that teachers did not do any better than chance at predicting which learning disabled students would gain the most from a read aloud accommodation on a mathematics achievement test. Of interest, however, Koran and Kopriva found that teachers could easily and clearly specify the different needs of students. What they appeared to struggle with was the assignment of accommodations to students based on these particular needs. This was echoed in a recent Plake and Impara report (2006) where educator experts of students with disabilities could understand differential needs of students across a broad spectrum of disabilities but had trouble linking these needs to particular accommodations. It seems that, while teachers may be proficient in identifying needs of individual students, at this time they have problems assigning differential testing accommodations based on these needs.

The focus group results reported by Douglas were revealing of the strategies ELL teachers use when given imprecise guidelines for making accommodations decisions for individual ELLs. In particular, teachers seem to work within the guidelines given to them, such as policies restricting the range of allowable accommodations, but, in general, they assumed that more accommodations were better. This is consistent with researchers' findings that teachers tend to over-assign accommodations for students with disabilities (Fuchs *et al*., 2000a, 200b) and that accommodations

decisions tend to be an "all or nothing" phenomenon (DeStefano *et al*., 2001). The common attitudes seemed to be "When in doubt, give the accommodation" and "If you are going to accommodate, give them everything you can." In large scale assessments, the focus groups suggested that teachers felt it was better to have a higher rate of false positives (giving accommodations to students who don't truly need them) so as to reduce the chance of false negatives (failing to assign an accommodation to a student who really needs it).

In general, present policies did not provide the level of explicit instruction necessary to allow teachers to make reliable decisions. Among the policies there appear to be few student-specific guidelines for assigning specific types of accommodations. As such, it does not appear that the policy-based guidance alone would provide appropriate guidance for making specific accommodations decisions for individual students who need them. It is hoped that, with more specific training and additional information to make good decisions, the rates of both false positives and false negatives could be decreased. An early computer-based system, the Minnesota Online Decision-Making Tool, used a policy-based approach for making accommodations decisions with students with disabilities and ELLs (Anderson and Spicuzza, n.d; Swierzbin *et al*., 1999). This model provided little support, however, for answering the questions of whether the student needed accommodations and which accommodations would meet that student's needs. It did attempt to systematize a process, though, that to-date had been given only scant attention. More recently, Elliott and Roach (2006) identified salient areas of need for students with disabilities and provided recommendations for focusing IEP teams on systematically addressing their diverse needs. Butler and Stevens (1997) developed theory-driven guidance for educators about how they might better approach the assignment of accommodations for ELLs based on linguistic and other student-specific variables. These will be discussed in some more detail in the next section.

Some authors attribute poor educator and possibly parent judgment with regard to assigning accommodations to inadequate training in the areas of measurement, standardized assessment, and accommodations (Hollenbeck *et al*., 1998). Indeed, DeStefano *et al*., (2001) found that their intensive and comprehensive teacher training program improved the quality and extent of accommodation decisions for students with disabilities. However, their results were confounded with two substantial changes in the state's testing program that may have also contributed to dramatic changes in accommodations decisions (Shriner and DeStefano, 2003). While it is important that teachers receive proper training in the areas of measurement, standardized assessment, and accommodations,

teacher training alone may not entirely solve the problem. Fuchs *et al*., (2000a) found that even when teachers read background information about test accommodations, discussed and reviewed this information with a research assistant, and had the opportunity to ask questions about the material, they still did not do a good job at assigning accommodations. Their work suggests that increased teacher education with regard to accommodations, by itself, may not be sufficient to improve accommodations decision-making.

In other areas of standardized testing, such as grading constructed responses and setting cut scores, human judgment is commonly and successfully used within highly structured, defined, and systematic approaches that involve routine oversight and auditing mechanisms. That is, detailed guidelines are given, judges are trained, and checks are put in place to assure the quality and consistency of the decisions. To not take a systematic approach to the assignment of accommodations seems to represent a breakdown of the systematized chain of evidence that leads to valid inferences (Mislevy, 1994). Such an approach to accommodations decision-making may help ensure not only the quality of the decision but also the consistency of the decisions across students, both of which are needed in order to make meaningful comparisons across schools, districts, and states. There has been a call for more systematized accommodation decision support systems for students with disabilities (for instance, Solomon *et al*., 2001). There are many critical variables to consider in each ELL student's background and many potential options for meeting the student's needs. A structured systematic approach is probably necessary in providing support for the decision makers who must take so much information into account in making the decisions, and more attention needs to be paid to making sure the essential decision-making elements are being included and fairly utilized.

Research-Based Approaches

There are a few possible research-based approaches for assigning accommodations to individual students based on their unique needs. Such systems are generally not intended to take the place of the persons who currently make accommodations decisions for ELLs and students with disabilities, but, rather, they are intended to provide a solid recommendation of accommodations which can be used as a basis for the person or team making accommodations decisions. In describing and discussing these approaches, we distinguish between those which are inductive and deductive in nature. Depending on purpose of the assignments, both foundations hold the promise of guiding high quality, consistent accommodations decisions. Each approach discussed below combines a data collection phase with an

accommodations assignment phase. Kopriva and others (Kopriva 2005a; Kopriva *et al*., 2006a, 206b) have recently emphasized that there appear to be four steps to developing a proper method for assigning accommodations, not just two. They report that key variables need to be identified, data need to be systematically collected, data from multiple sources needs to be thoughtfully combined in order to inform the assignment process, and, finally, a logic method of assignment that fairly and sensitively recognizes and matches salient student and accommodation characteristics needs to be built and implemented.

In most approaches (and especially in research-based approaches) where information must be combined to make an accommodations recommendation, building the system on a computer-based platform has been found to be quite valuable. Computerized decision support technology has been used successfully for many years in other fields, such as business, medicine, and law enforcement, for many years. For example, systems that produce differential diagnoses for medical personnel or track evidence patterns for law enforcement agencies effectively narrow down the routine work staff do. However, teachers and other team members are not asked to do this large-scale assessment assignment task on a regular basis so it is not surprising that their skills have not developed to produce reproducible decisions over like students or over teachers or districts. For this reason, systems that build research-based algorithms that effectively narrow choices would appear to be especially useful here. Perhaps resource limits have slowed down the development of such systems. It is only very recently that researchers have begun investigating ways to address this problem by utilizing findings and related literature. Most of these are computer-based because of the complexity of the work and the diversity of the populations they are focusing upon, and most of them are still in various stages of development.

Inductive Methods

In making accommodations decisions for ELLs and students with disabilities, it is possible to take an approach that is inductive, that is, to take specific incidences and make some sort of generalization. One reasonable inductive approach for assigning accommodations to individual students is the direct trial-and-error approach. In this approach different accommodations are tried with a student, one at a time or one package at a time, to see which ones help the student perform better in testing. This approach has been applied both informally and empirically in making accommodations decisions for individual students. In the focus groups with ELL educators, many of them stated that they used an informal trial and error approach to calibrate the appropriate accommodations for a specific

student by trying various accommodations in the classroom and tracking which accommodations worked best with that student (Douglas, 2004). Teachers felt they could use their observations to inform their accommodations decision making for large scale testing even in the face of a lack of clear, consistent guidelines for assigning specific accommodations to individual students.

Special education researchers have proposed and applied a formalized empirical version of this approach with students with disabilities. In this method the student is administered a short mini-test without accommodations and other parallel forms of the mini-test each administered with a different accommodation (Fuchs *et al*., 2000a, 2000b). The boost in performance on each accommodated mini-test over the performance on the unaccommodated mini-test is calculated and compared against normative data for regular education students who did not need accommodations. If the student's boost exceeds a certain cutoff in the normative data, then the student is deemed to need the accommodation. Their research found they could successfully predict for which students test scores on a large standardized test would improve substantially when administered with the specific accommodation or accommodations supported by the mini-test data. Recently, these researchers have published their operational approach which uses this methodology. It is known as DATA, the Dynamic Assessment of Test Accommodations (Fuchs *et al*., 2005) and is discussed briefly in the next section.

The formalized trial and error approach seems to effectively identify students who stand to benefit from test accommodations in research situations. However, limited work has been done to identify how well this works for large-scale testing purposes on a routine basis. Further, in contrast to a formalized trial and error approach, it has not been documented that informal teacher trial and error is effective in the same way. Other limitations need to be considered as well. First, this method is clearly time-consuming. For high school classes or other situations where the teachers or specialists work with many students this needs to be taken into account. Second, while Fuchs *et al*., (2000a, 2000b) found that their systematic formal approach was effective, small *n*s suggest caution. Third, unfortunately, the mini-testing can confound itself with some accommodation needs, such as the need for more frequent breaks. If the mini-tests are of shorter length, for instance, it may set up a different situation than what occurs during large-scale testing. Rather, especially in selecting large-scale test accommodations, conditions for the assignments need to be as close to those for which the accommodations are being considered. Several conditions, including test length and duration over days are usually constrained during this type of testing. Fourth, as with any approach,

students need enough time to practice any accommodation. This is true whether the students are preparing for a test or for the trial and error assignment process. Fifth, the pure trial and error approach typically does not account for interaction effects when multiple accommodations are needed, unless packages are tried out as well as individual accommodations.

Finally, constant trial and error may be a little like inventing the wheel over and over again. As the field determines that specific steps along the decision process are well-considered and sufficiently differentiated at the present time, it makes sense that systems can capitalize on this work and use the trial and error approach where certain distinctions or tasks are still in doubt.

Tindal and others (see below under *Deductive Methods*) are investigating the use of a hybrid form which combines trial and error and other methods. These and others will be considered next.

Deductive Methods

Deductive approaches based on theory, or generalized notions which have been found to hold promise, also have been proposed. Theory-driven deductive approaches focus on identifying critical factors that are related to the individual needs of students and the active characteristics of test accommodations. Abedi (2007) emphasizes that only a small subset of accommodations is most appropriate for ELLs. While he didn't specify which factors in these accommodations were relevant or for whom, other researchers have addressed this issue (Kopriva *et al*., 2007a). In reviewing the ELL literature for key student indicators, the researchers who developed the Selection Taxonomy for English Language Learner Accommodations (STELLA) system discussed below compiled a list of 139 different variables important for learning and evaluation in general. For the particular purpose of large-scale testing accommodations, however, their analysis found that only a small number of these seemed to be salient for this type of situation, and these were discussed in the first section of this chapter. The extensive review was useful, however, because it helped to identify some nuances not generally considered in large-scale assignment to date, and provided information which could be in the development of the logic or research-driven systems. For example, August (August *et al*., 2004) reported that access to bilingual glossaries have been found to be effective for understanding test requirements in English when ELL students have at least some literacy in their home language. This is because they could transfer cognate cues across languages. The students did not need to be fully literate in their home language to begin using this skill; however, the more literate they were in their native language the more they could avail themselves of this benefit. Her work provided developers of the system guidance about

how to assign this accommodation and to whom. Below, theory-driven approaches have been classified as those that focus on a direct assessment of ancillary skills and those which use an informant approach.

Direct Assessment of Ancillary Skills This approach attempts to directly assess selected ancillary skills of individual students that are unrelated to what the test is intending to measure but may block the student's access to the test if he or she does not have a sufficient command of this skill. Skills related to what the test is intended to measure may also be used provided they are predictive of what accommodations would help the student overcome access barriers in taking the test. Researchers interested in improving accommodations decisions for students with disabilities have attempted to use measures of ancillary skills, in particular reading, to predict whether those students would benefit from a specific accommodation (Helwig and Tindal, 2003). Tindal and others have also included measures of these skills in more comprehensive systems, such as their recent method, the Accommodation Station, which is identifying and evaluating the impact of selected student data related to accommodations and what that might mean for accommodations assignment (Ketterlin-Geller, 2003; Tindal, 2006). Directly measuring the level of access skills within the assignment system may be helpful in selecting individual accommodations for ELLs, particularly because some of the skills clearly impact student access to large scale tests in other content areas as they are currently conceived.

Like the trial-and-error approach, the direct assessment of ancillary skills involves testing the student directly. Thus, the measurement of the particular ancillary skills must also be accurate to get valid results, and this, in itself, is a large undertaking. For instance, much literature and many testing products illustrate how complex and resource consuming it is to effectively develop measurements of reading proficiency. For those who assess ancillary skills, specific domain and theory-driven variables need to be identified and measured, and logic about how to combine these to obtain proficiency levels would need to occur. Initial attempts at this approach that used an assignment system to making accommodations decisions for students with disabilities were largely unsuccessful (Ketterlin-Geller, 2003). A current project is attempting to build on this work by beginning to identify decision-making logic partially informed by the ancillary skills related to reading that the latest version of the Accommodation Station provides (Siskind *et al*., 2004; Tindal, 2006). This ongoing work will be summarized in the next section.

While Tindal *et al*.'s method utilizes a direct measure of some skills, they also recognize and have attempted to incorporate other "informant data" in this system. Helwig and Tindal (2003) used reading and math

pretest scores to try to build student profiles that might better predict whether students with disabilities would benefit from having the read aloud accommodation. However, the results were often in conflict and their system did not perform much better than teacher judgment. They suggested that their analysis failed to take into account a potentially critical variable, namely the linguistic complexity of the items in their test of ancillary skills. This complexity is widely recognized in the ELL literature (for instance, see Kopriva, 2000; Abedi and Lord, 2001; Abedi *et al*., 2000, 2001). Student performance with accommodations related to linguistics can be very complicated behavior to model and predict because of possible interactions with the characteristics of the test items. As such, data on the language complexity of the large scale assessment the student will be using seems to be necessary in determining the effectiveness of some accommodations, or test forms with a identified levels of linguistic complexity must be recommended as one of the accommodations for particular students.

Tindal *et al*.'s current system uses some of the Fuchs and Fuchs trial and error methodology as well in that they test students with and without a read aloud accommodation. However, recent data related to incorporating this methodology with his other work are inconclusive and disappointing. Finally, Helwig and Tindal (2003) speculated that their system was too simplistic and failed to benefit from the teachers' intimate knowledge of their own students. They proposed that math and reading screening test scores be utilized in tandem with teacher knowledge of individual students might produce better accommodations decisions. Lately, researchers of the Accommodation Station have attempted to address this limitation by including questionnaire data from teachers and the students (Tindal, 2006). This method is also employed by other researchers and will be discussed next. To date, it is unclear how data from the direct assessment of ancillary skills, the informant data, and the trial and error data will be combined in the Accommodation Station to provide guidance to IEP teams. Structuring a system that incorporates different methods, however, is interesting, and it is anticipated that future work will focus on these next steps.

Informant Approach The informant approach relies directly on the insights of those who know the student personally and/or other extant skill information that would be available in the student's record. In this sense, it has benefits over the other approaches discussed so far because is not limited to skills or needs that can be tested or measured directly. In fact, it may not be necessary for the student to be directly involved in the process at all. The data are indirectly collected, however, which introduces another source of error and caution. Using this approach, information about the student may be collected from one source only (for instance, the teacher)

or gathered from multiple sources, such as the student, the student's teacher, and the student's parent or guardian. As defined here, specific information is gathered using standard instruments specifically designed for the purpose of assigning large-scale test accommodations. Further, the intent of this method is to focus instrument questions on collecting clearly delineated information about critical variables that the developers have identified as being salient for assigning accommodations to individual students. Therefore, the informant approach is distinct from the policy-based teacher or team approaches in that method does not rely solely on instinct or sources of information about students and accommodations that are vaguely and implicitly utilized and combined in a non-standard way. On the other hand, depending on the informed source(s) and whether it is collecting data obtained from other test(s) or relying on the judgment of adult(s) who know the student well, the informant approach may rely on information, systematically collected, that is skewed (human judgment of parents or teachers) or outdated (for instance, information in the student record). The quality of the student data which is collected relies on the quality of the information gained from the informants. Further, the quality of the decision is to some extent dependent on the quality of the informant's knowledge and the acumen of the researchers who interpret theory and build the algorithms. Like any approach, the quality also is bound to be variable for different students.

No literature has been found which systematically links only student records information and accommodations assignments for large scale assessments for students with disabilities or ELLs. Likewise, no research has been reported which combines uniformly collected information about student needs collected from only one "human" source (for instance teachers or parents) with this one human source making assignments based on this information. As noted above, Koran and Kopriva (2006) reported results when ELL teachers, by themselves, were asked to assign accommodations based on a standard collection of data from multiple sources, including themselves. Even when standard instruments were used, teacher assignments proved to be no better than a random assignment of accommodations for individual students.

STELLA, the Selection Taxonomy for English Language Learner Accommodations, is a newly developed informant system designed to assign individual accommodations for K-12 English language learners (Kopriva *et al*., 2006b, 2007a). Currently, it utilizes findings from the multiple sources (parents, teachers and records) and systematic student data collection procedures. It identifies critical variables, collects data, combines the data with standard information regarding how accommodations perform, and then uses a standard series of computerized algorithms based in theory

and formative empirical input. The system is built to utilize the latest information about students that appear to be the most relevant for making accommodations decisions about this population. Further, it is designed to be customized to accommodate the policies of different states or districts while also retaining the ability to provide a systematic series of recommendations over agencies which have been found to be meaningful based on best practice and empirical findings. Two initial verification studies found that this system seems to be producing appropriate decisions for individual students. Future work will continue to examine and refine the extensive series of algorithms, update the preloaded aspects of STELLA and possibly add a direct component to the system. This system will be discussed in more detail in the next section.

It is argued that, when direct testing is utilized or multiple sources are consulted, the research-based approaches may mark an improvement over pure policy-based approaches not only by formalizing the content and contribution from each source, but also by separating the role of student advocate from that of primary decision maker. Douglas concluded that ELL teachers felt, in the present policy-based climate, that they were asked to simultaneously take on competing roles as both expert decision maker and student advocate (2004). This may present a dilemma for teachers who are asked to make decisions with minimal guidelines and support while maintaining the best interests of the child. Two sets of work (Koran and Kopriva, 2006; Plake and Impara, 2006) have found that educators struggled to identify proper accommodations even though they could easily identify needs of individual students. Perhaps educators may be better at some roles than others in this accommodation assignment process. Each of the systems summarized below utilizes research and best practice as a foundation of helping educators narrow down choices and provide guidance. As such, they are designed to improve the selection approaches necessary for making informed decisions about accommodations in research and in practice.

Operational and Prototypical Systems

Within the last 10 years there have been some attempts to create tools for improved accommodation decision-making that are individual student based. Research and other information associated with these systems have been mentioned above. This section will explain the methods in somewhat more detail. Three guidance models, one for both students with disabilities and ELLs, one for students with disabilities, and one for ELL students, will be discussed first. Next, three systematized research-based systems, two for students with disabilities and one for English language learners, will be

summarized. The goal of all of these methodologies is to show improvement over current practice, which is used as the comparison.

Guidance Models

Following are three examples of guidance models. Two of the three are hybrids of policy-support and informant research based approaches, while the first system which will be discussed used the policy-support approach only.

MINNESOTA ONLINE DECISION-MAKING TOOL

An early system, computer-based, used a policy-based approach for making accommodations decisions with students with disabilities and ELLs (Anderson and Spicuzza, n.d; Liu *et al*., 1999; Swierzbin *et al*., 1999). This tool attempted to improve accommodations decision-making by guiding parents, administrators and others through a decision tree based on one education agency's test accommodation policies, and was created on the basis of existing state guidelines for the inclusion of these students. However, it did not collect data to make the decisions at the various points. Rather, the users were responsible for collecting the relevant information and making a correct decision at each point. They were expected to address the requirements at each step by relying on their own knowledge of the student and knowledge of accommodations. Feedback from ELL/bilingual educators, coordinators, and administrators reviewing the tool for use with ELLs indicate that they struggle with obtaining the type of knowledge needed to answer the questions posed by the tool at some of the decision points (Liu *et al*., 1999).

Specifically, the system directed the user through the decision tree to determine whether policy dictates that the student should be exempted, (eligible to) participate with accommodations, or participate without accommodations. The tool then pointed each user to specific policy information, such as what accommodations are available and the tests for which they are appropriate. However, it left the decision about assignment of specific accommodation(s) for specific students to the user. Finally, the tool also facilitated the implementation of accommodations by providing information about whether materials need to be special-ordered in order to provide the accommodation to the student for the state assessment. The developers emphasized that one of the advantages of the online system is that updates to policies on the online system would be easier for educators to manage than updates to printed policies. This assured that teachers are using the most current version of state accommodation policies to make accommodations decisions.

This model provided little support for answering the questions of whether the student needed accommodations and which accommodations would meet that student's needs. It did attempt to systematize a process, though, that to-date had been given only scant attention. A pilot of this system found mixed initial acceptance among IEP team members who used the system in making accommodations decisions for students with disabilities (Swierzbin *et al*., 1999). However, a pilot with ELL/bilingual educators, coordinators, and administrators found a more positive response for the use of the system with ELLs (Liu *et al*., 1999). A system such as this may improve the implementation of educational policy, but appears to offer little specific guidance to the user who must decide which accommodations are most appropriate for an individual student.

Five-Step Systematic Decision-Making and Documentation Process

Stephen Elliott and others (e.g., Elliott *et al*., 1999; Elliott and Roach, 2006) have continued to provide guidance to IEP teams about how to wisely assign large scale accommodations for students with disabilities. In the *Assessment Accommodation Guide* (Elliot *et al*., 1999), *Assessing One & All . . .* (Elliot *et al*., 2001), and other work, these researchers have identified key information and student needs that teams should know, critical access skills that are particularly salient for this population, and process factors that influence accommodation decision-making. The *Assessment Accommodations Checklist* and associated guidance directs IEP team members through the accommodation selection, implementation planning and documentation processes. The authors encourage members to link any of the sixteen key access skills they have identified as being problematic for an individual student to one or more accommodations that specifically minimize interference between conditions and measurement of target skills. These skills represent elements of typical large-scale standardized testing conditions which could pose a problem for students with disabilities. Some of these are also problematic for English learners as well. The sixteen access skills are:

- Attending
- Listening
- Reading*
- Remembering
- Writing
- Following directions
- Working alone
- Sitting quietly

- Turning pages
- Locating test items
- Locating answer spaces
- Erasing completely
- Seeing
- Processing information in a timely manner
- Working for a sustained period of time
- Spelling.*

* Depending on what is being measured these could be target skills and, if so, they should not be accommodated.

Butler and Stevens Guidance

In 1997 Butler and Stevens completed a report for CRESST (Center for Research on Evaluation, Standards, and Student Testing). This report provided an overview of ELL student characteristics in general and guidance about which factors are particularly important in selecting appropriate accommodations to individual students. They also provided some direction about how to use this information. While they didn't address this report to agencies or teachers *per se*, the guidance would be helpful to those making accommodation decisions for students.

The three student factors these authors felt were critical were English language proficiency, prior formal schooling, and length of time in U.S. As they detailed, it is important that appropriate tests of English language proficiency include the measurement of complex academic language, conceptual and discourse skills, not just factual academic vocabulary and knowledge levels. The recent generation of these tests is improving the ability to measure these advanced and more complex dynamics of English proficiency as well as more basic levels of English literacy and fluency. They appear to be closer to what Butler and Stevens were recommending. They felt that the second variable, prior formal schooling, was critical because students with little or no formal schooling tended to have less experience with schooling and testing conventions, especially those used in the U.S. Further, they recognized that not all formal schooling was similar; in some countries the knowledge, skills and methods were quite distinct from U.S. schools. Therefore, it was important to identify both amount of formal schooling and the kind of formal schooling the students had experienced. Those with significant formal schooling in their home countries are more likely to be literate in their L1 as well, and this suggests that students will move more quickly from accommodations to no accommodations as they become more proficient in English. It also widens the types of accommodations which may be appropriate. The third variable, length of time in the United States, was chosen because it suggests exposure to

U.S. schooling and testing conventions and the U.S. mainstream culture in general. They acknowledge, however, that isolation of home communities from mainstream expectations is an important covariate that should be considered.

Figure 10.5, from their report, suggests how an agency might use this information. Unfortunately, however, it only provides guidance for three choices: no accommodations needed, accommodations needed, or exemption preferred. The report also includes a background questionnaire for middle school students to complete which can elicit information about the variables they have identified. The ten questions include where and when the student was born, how long they have lived in the U.S., if they went to school in another country and, if so, where and how long. It asks what is their first language, what is spoken in the home and whom do they speak it with, if they ever had formal study in a language other than English and if so, what language and for how long and where did they study (home, school, language school or other). Finally, the questionnaire asked if they have ever learned content in another language (and if so, what), how many years have they attended school in the U.S, and if they are have learned/are learning content in the U.S. and if so, which subjects are they studying.

While this report does not guide decision-makers to specific accommodations for individual students it does provide a view into the types of data which would be needed to make those decisions. As such it represents an early attempt at categorization that the *STELLA* system (below) has built upon.

Research-Based Standardized Data Collection and Assignment Systems

Three systems will be outlined here. In general, the development of the deductive research-based systems has lagged behind that of inductive methods, in part because of the technological requirements and development costs associated with them.

DATA

DATA, the Dynamic Assessment of Test Accommodations (Fuchs *et al*., 2005), is a systematic method for assigning test accommodations to students with disabilities in grades 2–7. DATA uses a pure empirical trial-and-error approach as described earlier in Fuchs *et al*., (2000a, 2000b). In this method the student is administered a short mini-test, without accommodations and other parallel forms of the mini-test each administered with a different accommodation. The boost in performance on each accommodated mini-test over the performance on the unaccommodated mini-test is calculated and compared against normative data for regular

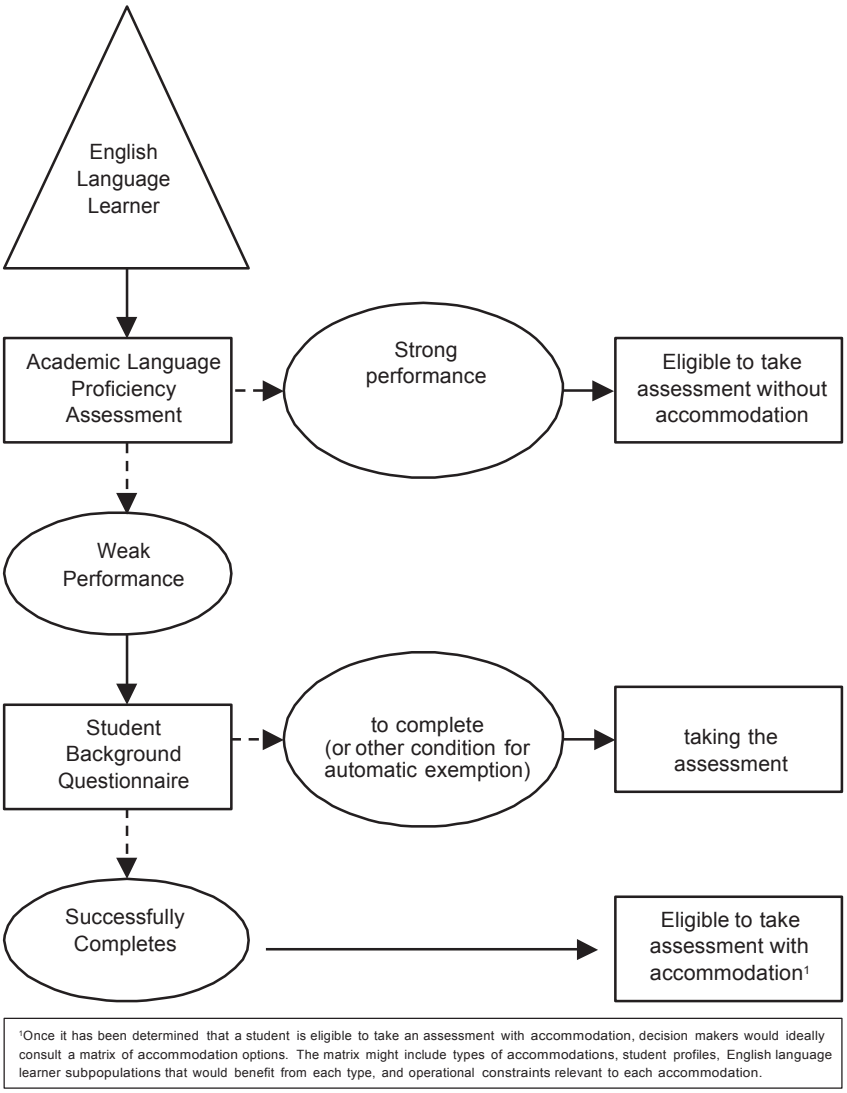

Figure 10.5 Potential Screening Process for Accommodation Eligibility Decisions. Adapted from Butler and Stevens, 1997.

education students who did not need accommodations. If the student's boost exceeds a certain cutoff in the normative data, then the student is deemed to need the accommodation. DATA assesses student need for accommodations in tests of reading and mathematics. Research has indicated that the system is successful in predicting which students will

benefit from specific accommodations on full length standardized tests of math and reading (Fuchs *et al*., 2000a, 2000b). Unlike the next two systems we will discuss, DATA is now a fully operational product published by PsychCorp (Fuchs *et al*., 2005).

Accommodation Station

The Accommodation Station is a computer-based system currently being developed by researchers at the University of Oregon which is primarily a tool for students with disabilities (Tindal and others, e.g., Tindal ,2006; Tindal and Fuchs 1999; Tindal and Ketterlin-Geller, 2004). Rather than being available now or soon, the Accommodation Station is an approach around which a flurry of related on-going research studies were and are being conducted. The central large-picture concept of the system is focused on examining aspects of assignment for future computer-based systems that will not only provide information about accommodation assignment but also deliver the accommodations on line, and test students in the content areas of interest (Ketterlin-Geller, 2003). Early studies in this line of research focused on directly assessing student access skills (such as reading and math skills) and using these scores to predict whether students would benefit from certain test accommodations (Helwig *et al*., 1999; Helwig and Tindal, 2003). More recent studies have added student and teacher questionnaires (Alonzo *et al*., 2004; Tindal, 2006) and have continued to investigate student characteristics that predict whether a student will benefit from a particular test accommodation (Tindal and Glasgow, 2005).

Virtually all of the research to-date has been done on students with disabilities although there has been movement lately associated with adding Spanish proficiency as an ancillary skill to be measured at some point. A trial and error approach similar to Fuchs *et al*.,'s *DATA* system has also been incorporated into a recent prototype of the Accommodation Station. Student performance on a subset of math items in read-aloud and self-read conditions are compared. Unlike, the Fuchs and Fuchs method, boost is assessed under these two conditions for a student rather than compared against normative data as in the *DATA* system. Early research flagged a number of problems with using a purely direct approach though (Helwig *et al*., 1999; Helwig and Tindal, 2003). The Accommodation Station has now added an informant aspect although initial forms and data being collected in this way have been problematic (Tindal 2006). Its expert-driven decision-making algorithms, focused on students with disabilities only, are in development (Siskind *et al*., 2004; Tindal, 2006).

Variables considered in recommending accommodations for individual students include the results from the direct method of assessing ancillary information, such as the ability to use a mouse to enter responses on a

computerized test, reading comprehension, and level of math skills when these are considered to be ancillary in nature. It also includes input from the student on his or her preferences for accommodations and teacher's input on factors such as classroom use of accommodations. True to its direct approach roots, data collection in the Accommodation Station is centered on the involvement of the student, both in measuring the student's ancillary performance and in using the student as a knowledgeable informant of the types of accommodations that would be helpful. Recent findings suggest, however, that the student is remarkably inconsistent with what they perceive to be useful accommodations for themselves (Tindal, 2006). Tindal also reports that some teacher information appears to be inconsistent as well.

Beginning in 2005 an expert panel of participating special education state specialists was convened. A systematic consensus-building method for assigning accommodations to different kinds of students using a case-study approach was employed in late 2005 and 2006. A final report (Plake and Impara, 2006) explained the progress in utilizing such a method and provided guidance about how to continue towards the assignment goal. It is anticipated that future funding will augment the work so far and subsequently complete this part of the Accommodation Station's development.

STELLA

While some attempts had been made to look at how to match students with disabilities with appropriate test accommodations, very little prior research had been conducted to look into the testing needs of individual English language learners and establish systems for matching ELLs with appropriate test accommodations. The Selection Taxonomy for English Language Learner Accommodations, STELLA, has been developed over the last three years and is in a draft final form (Kopriva, 2002, 2005a, 2005b; Kopriva and Hedgspeth, 2005; Kopriva and Mislery, 2005; Kopriva *et al*., 2006a, 2006b). STELLA is intended to be used with K-12 students to assign accommodations to the range of ELL students for use on large-scale academic assessments. It is a computerized system, research-based, and empirically developed and verified. It has been built, most recently, from research conducted as part of two large, multi-year projects. As more evaluation of the system is completed, it is expected that states or other interested users will soon be able to utilize the "beta" version of the system.

STELLA utilizes the informant approach to making test accommodations decisions for individual ELLs. It has been built on a flexible platform to adjust to the needs and obligations of different users, and has been designed to be compatible with and flexible enough to accommodate information from the direct approach if future research suggests this is a

viable addition to the system. At the present time this system uses three forms to collect the most salient data about the needs and strengths of each student from records, parents, and teachers. The records form is based on information that is in the student's file at the school, such as English language and native language proficiency test scores and other information about the student's schooling in the U.S. The parent/guardian form is based on an interview with the student's parent or guardian regarding the student's proficiency in the home language, prior schooling opportunities, and prior experiences with classroom and standardized testing. The teacher form is based on observations the teacher has made about the student's language proficiencies on an ongoing basis in the classroom as well as their apparent preferences and accommodation experiences in the classroom. In the STELLA system teachers are not asked to recommend large scale accommodations at the present time as the authors suggest that research says this currently appears to be difficult for these educators to do.

These data are compiled and then subjected to an extensive set of conversion, consolidation and decision-making algorithms. Output includes an individualized student profile and recommended accommodations for each student. State or district-allowed accommodations that are recommended are highlighted but additional STELLA recommended accommodations are also shown. This allows users to utilize the system under current legislation and regulations and provides some guidance about how these policies may be improved upon for individual students. As noted above, the system is developed so that highlighted accommodations change by educational agency, and decision-making algorithms can be added to allow for additional accommodations. At this time specific accommodations that research and practice have suggested are promising for this population have been preloaded into the system. Likewise, since part of the system is designed to convert different proficiency test scores to common scales, certain tests are preloaded and others can be added as needed.

Another project[1] offered some early insights into salient student and accommodation characteristics and how a matching model using an informant approach might work as an operational system. In this project some relatively simple matching systems were developed, each designed exclusively for the specific accommodations study at hand (Kopriva, 2005a; Kopriva *et al*., in press; Emick and Kopriva, 2006). Since the focus of this project was elsewhere, a later project built on this work and focused exclusively on developing a rigorous assignment system for English language learners. Once funding for this second project was retained, the forms were conceptualized, designed and developed after extensive developmental work that reviewed current large-scale and instructional accommodation literature, conducted focus groups, obtained recommendations from

targeted practitioners, and subsequently narrowed down the most salient student variables. Key variables were selected and interviews with teachers and parents and feedback from partner educational agency staff refined the questions on the forms. Concurrently, project staff, with guidance from educational agency participants, built on knowledge gleaned from the earlier project, identified the most promising accommodations, and completed an evaluation of the active characteristics of the final set of accommodations. Algorithms were developed by project staff and consultants and reviewed and edited by an expert panel review (see Kopriva, 2005a; Kopriva *et al*., 2006, 2007a, for more of an explanation of the formative studies and related development work).

In its present form STELLA does not involve the student directly. Thus, it can be used with students who are quite young. The accommodations matching system initially included a trial and error component that asks the teacher(s) who have worked with the student in the classroom which accommodations the student needs. However, this information is not utilized at this point in the STELLA decision-making, but was included for research purposes only. Unfortunately, Koran and Kopriva (2006) found that teachers' understanding of student needs was no different than random and so this part will probably be deleted from the system. As mentioned above, the system utilizes direct skills information to the extent that it includes test score data in the decision-making algorithms, and considerations in the formative algorithms that flag data which are too dated. Both the trial and error information could be more formalized and incorporated with the informant approach. STELLA takes into account student performance on previous standardized testing, the ELL's experiences with testing accommodations in the classroom, the teacher's own intimate knowledge of each student's needs and preferences as evidenced in the classroom, as well as many other variables that contribute to the recommendation of a specific package of accommodations for an individual ELL at a given point in his or her learning.

Once STELLA was completed, two initial validation studies were conducted in 2005. One study collected data on the feasibility of the system (Kopriva *et al*., 2007). It also researched how the STELLA findings compared with teacher recommended assignments (Koran and Kopriva, 2006). Nineteen teachers who spanned the range of K-12 participated from three states, each selecting six beginner to advanced ELL students. A total of 114 sets of files were completed with each set comprising three completed forms per student. Feasibility results showed that teachers, who were the coordinators of the data collection on each student, could clearly identify needs of diverse sets of ELL students across schools and states. Findings suggested that the system collected consistent data from like questions over

forms, to be used for triangulation and confirmatory purposes, and that each form also appeared to contribute unique data about each student from the unique vantage point of each source.

Koran and Kopriva reported that, in addition to completing the STELLA protocols for their students, teachers were also asked to provide their own accommodation recommendations for each student at three points during the form completion process: before the forms were completed, directly after the data collection and based only on information STELLA had collected, and then later with any additional information they considered to be important about each student (but which was not represented on the STELLA forms). Subsequently, a blind panel of three ELL practitioners and an ELL researcher were convened and asked to independently rate how each of the teacher accommodation recommendations, the STELLA recommendations, and a randomly generated set of accommodations compared. These raters did not know the students; rather, they made their ratings after reviewing data from the forms for each of the 114 students. Rater findings were subjected to goodness of fit analyses. Results indicated that STELLA was found to consistently and significantly be the best fit, over students and over all recommendation alternatives. On the other hand, however, teacher recommendations disappointedly did not prove to be different as teachers learned more about each student from the data collection process. Nor were any of these teacher recommendations significantly different from the random assignment.

The purpose of the second study was to investigate if students who received selected STELLA recommended accommodations performed better on the test relative to those who received improper accommodations or no accommodations whatsoever (Kopriva, Emick *et al*., in press; Kopriva *et al*., 2007). It was designed to provide information about the validity and effectiveness of the particular STELLA assignments utilized in the study. Initially, 276 third and fourth grade South Carolina ELL students who spanned the range of English language proficiency completed a computerized mathematics test under randomly assigned accommodations that were implemented electronically as the students took the test. Three accommodations were used (oral English, bilingual word translation, and picture-word "translation") and students randomly received one, two or three of them. One group received no accommodations. Afterwards, additional data about the students was used to assign students to one of the three groups (proper accommodations, improper, or no) as per the STELLA framework. Findings showed that students who received proper accommodations performed significantly better than either students receiving improper accommodations or no accommodations. It also showed that students who received inappropriate accommodations (as per

the STELLA assignment) scored no better than those who received no accommodations. This study not only verifies the reasonableness of the affected STELLA assignments but also suggests how important it is for students to receive proper accommodations vs. improper ones.

It is anticipated that future work with STELLA will continue to investigate its usefulness and effectiveness. Developers are interested in studying if the system works well for ELL students of all backgrounds and at all proficiency and grade levels.

Implications

At this point it should be clear that options in the ways test materials are presented and administered is very important for students who come from diverse cultural backgrounds and have English literacy and language challenges. However, no matter how adroitly and carefully these options are selected and created, if they don't get to the correct students who need them, they are not useful. Emerging research suggests that use of improper accommodations may result in scores that are not substantially more valid for these students than if they received no accommodations. This is very troubling for many reasons but particularly two. First, individual performance on large scale academic tests and scores may misrepresent the skills of students who do not receive proper accommodations. This can have profound consequences for future learning opportunities for these students, including improper placement or other student level decisions, as well as inappropriately impacting educational agency level decisions and possibly resource allocations. Second, improper accommodations may help explain why the findings from accommodations research have, to date, been so mixed. It may be suggested that only when researchers employ rigorous means of matching students to accommodations they are empirically testing will their work provide clearer results about which options are useful and which are not. Currently, many of us are left with the vague notion of which accommodations are helpful but we're not sure how or when. The real question may well be "Useful for whom?" It is argued here that this is the question researchers should be investigating.

Many student factors are critical in understanding how to properly instruct ELL students as they become more proficient with the language and conventions of U.S. schooling. This chapter suggests that a subset of these key indicators appear to be salient in determining which large-scale testing options a student should receive, especially as long as standard assessments of academic content remain similar to those which are currently in use today. Test purposes, environments, standardized conditions that assume few if any procedural or material adaptations, and reliance

on written language (most typically in English) mean that some student issues as compared with others take on more importance in the current testing climate. Several of these have been identified here and, it is argued, should be used to identify and determine accommodation need for individual students. Likewise, certain accommodations or testing options should be clearly highlighted as being especially useful for ELL students with certain needs or strengths. These have been identified in the previous chapters. In tandem, the recognition of these characteristics appears to set the stage for developing ELL matching systems which can be useful for both practitioners and researchers.

To date, the form that these systems should take is not entirely clear. Simple policy guidelines do not appear to be sufficient. It is possible more specific guidance may be helpful and effective, with some type of systematic training and rigorous oversight procedures in place. Certainly, teachers often know the students well, and IEP teams utilized with students with disabilities suggest that other personnel or even the students themselves provide important and unique insights. Relying exclusively on the knowledge and time associated with compiling materials, educating and utilizing overtaxed educators, parents and students, however, may not be practical or reasonable even though in less demanding circumstances they may be "up to the task". Also, research suggests teachers or other educational specialists are particularly well positioned to identify needs but, to date, appear to struggle with how to assign accommodations. This may or may not extend to parents, other school personnel and students. Computerized systems that guide personnel through the policy process provide standardization and, in the future, may also provide selected educational opportunities as the needs arise within more complex programs.

Electronic systems that provide differential boost information or that directly test the levels of students' ancillary skills are interesting because they provide up to date needs information that offsets records or other data which may be dated or educator judgments that may be skewed. However, to date, they appear to be time consuming for students. Informant approaches should help streamline the process; however, unless stringent evaluation is part of development and ongoing implementation, they may provide a too-simplistic set of recommendations which would not properly address the nuanced and ever-changing needs of these populations. All in all, it appears that theory-driven systems need to be computerized as long as the algorithms appropriately address the complex realities associated with accommodations assignment.

As system designs mature, any of them will need to have the ability to handle relevant data in a way that sensitive decisions can be made not only for each student but consistently and even-handedly across students. It

is also anticipated that most future systems will be hybrids in some form or another. Such systems will need to have the capacity to prompt, include and combine large numbers of discrete pieces of information in complicated ways with ease and be able to update frequently and store algorithms and output for use by different educational agencies. All in all, methodologies would probably represent variations of expert (or knowledge-based) systems used in other disciplines (for instance see Wright and Bolger, 1992) where data collection sources include humans as well as other knowledge-based data about students and characteristics of testing practices.

Besides continuing to develop more sophisticated data collection methodologies, continuing attention to the algorithms that are used to convert and combine data and to the decision-making rules which match student factors with accommodation characteristics will be important. Additionally, there should be more investigation of the teachers' role as student advocates and how this affects their input in systems. The effect of stakes on the accuracy of teachers' responses on questionnaires and the effect of accountability in the decision making system on teachers' judgments may be useful areas of inquiry as well. Third, implementation issues need to be identified and evaluated. There is currently little published work on these implementation considerations of utilizing these matching systems but there are likely to be practical issues associated with them. For instance, optimal timing to use the systems in order to provide a recommendation should be considered. Accommodations decisions must be made far enough in advance to allow for any practice and to allow sufficient time to set up the logistics of implementing the accommodation for the large scale testing sessions. Relative to many students with disabilities, the needs and abilities of English learners are constantly and quickly changing as they become familiar with U.S. culture and testing practices. Research should consider what amount of time before testing would allow for the accurate assessment of student ability and need while still allowing sufficient time to implement the recommendations appropriately. Further, while more attention is being given to getting the right accommodations options matched with the right students based on individual student need, the next link in a valid inferential chain is making sure that the accommodations actually get administered to the student, and get administered properly. Some limited research literature (e.g., Jayanthi, *et al*., 1996; Solomon *et al*., 2001) and extensive practitioner experiences suggest that there appear to be issues with implementing test accommodations that have been recommended through current operational processes.

Finally, some researchers envision accommodation assignment approaches that are integrated into computerized content area testing

systems. Kopriva *et al*., (2006b) discuss future testing systems that are sensitive to a number of nuances about the ELL students and alter accommodations by item. These systems would not only allow for matching algorithms to operate in them, such as that produced by STELLA, but they would include algorithms that address item characteristic data which are more finely tuned to student needs. For instance, Solano-Flores *et al*., (2002) suggest that ELL students differentially use native language or English items when presented with dual language forms. While others (e.g., Duncan *et al*., 2005) suggest students use primarily one form or another, Solano-Flores suggests that his more fine-grain research points to small preferences depending on a number of linguistic factors. In a similar fashion, Ketterlin-Geller (2003) predicts that Accommodation Station may one day implement recommended accommodations for students with disabilities directly within a computer-administered academic test system. That is, similar to the thinking of Kopriva and others, she suggests that the accommodations matching system could be embedded in a larger assessment system that also administers the accommodations

Endnote

1. The Valid Assessment of English Language Learners (VAELL) project at C-SAVE at the University of Maryland.